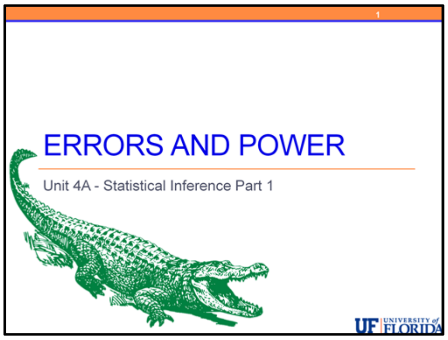
ERRORS AND POWER

Unit 4A - Statistical Inference Part 1

UF UNIVERSITY of FLORIDA

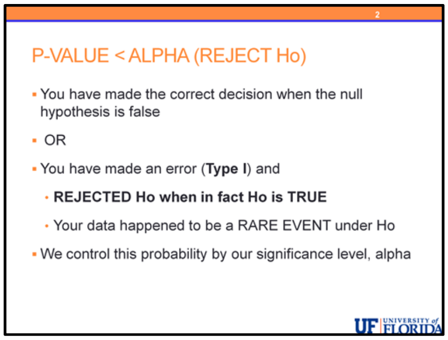


We are not guaranteed to make the correct decision by this process of hypothesis testing.

Maybe you are beginning to see that there is always some level of uncertainty in statistics.

P-VALUE < ALPHA (REJECT Ho)

- You have made the correct decision when the null hypothesis is false
- OR
- You have made an error (**Type I**) and
  - **REJECTED Ho when in fact Ho is TRUE**
  - Your data happened to be a RARE EVENT under Ho
- We control this probability by our significance level, alpha



When we conduct a hypothesis test, we choose one of two possible conclusions based upon our data.

When the p-value is less than alpha we reject the null hypothesis and either:

You have made the correct decision when the null hypothesis is false

OR

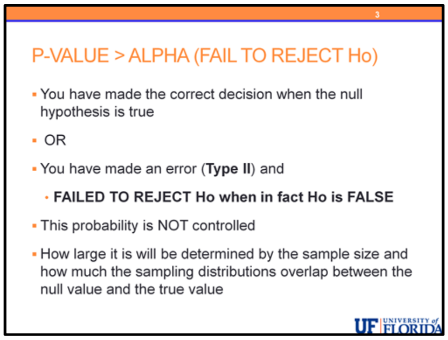
You have made an error (called a **Type I error**) and

- **REJECTED Ho when in fact Ho is TRUE**
- In this case, your data happened to be a RARE EVENT under Ho

We do control this probability by our significance level, alpha and thus in general this will happen no more than 5% of the time using the common significance level of 0.05.

P-VALUE > ALPHA (FAIL TO REJECT Ho)

- You have made the correct decision when the null hypothesis is true
- OR
- You have made an error (**Type II**) and
  - **FAILED TO REJECT Ho when in fact Ho is FALSE**
- This probability is NOT controlled
- How large it is will be determined by the sample size and how much the sampling distributions overlap between the null value and the true value



When the p-value is greater than alpha, we fail to reject the null hypothesis and either:

You have made the correct decision when the null hypothesis is true

OR

You have made an error (called a **Type II error**) and

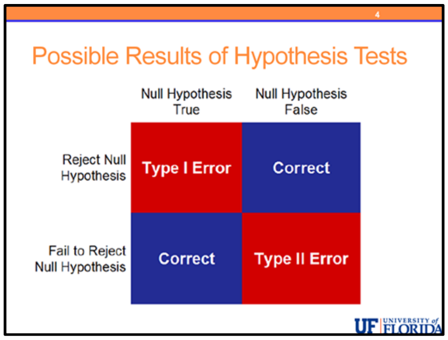
- **FAILED TO REJECT Ho when in fact Ho is FALSE**

This probability is NOT controlled.

How large it is will be determined by the sample size and how much the sampling distributions overlap between the null value and the true value.

Possible Results of Hypothesis Tests

	Null Hypothesis True	Null Hypothesis False
Reject Null Hypothesis	**Type I Error**	**Correct**
Fail to Reject Null Hypothesis	**Correct**	**Type II Error**



Here is a visual representation of the four possible choices in hypothesis testing.

No matter which decision we make (reject the null or fail to reject the null) there is the possibility of making either a correct decision or an error.

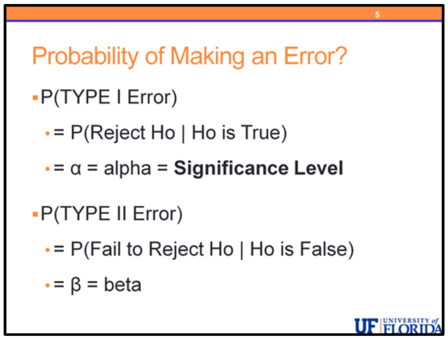
A Type I error occurs when we falsely reject a true null hypothesis and

A Type II error occurs when we fail to reject a false null hypothesis.

Be sure to keep these two errors straight. This image can help.

Probability of Making an Error?

- P(TYPE I Error)
  - = P(Reject Ho | Ho is True)
  - = α = alpha = **Significance Level**
- P(TYPE II Error)
  - = P(Fail to Reject Ho | Ho is False)
  - = β = beta



Both of these probabilities are conditional probabilities.

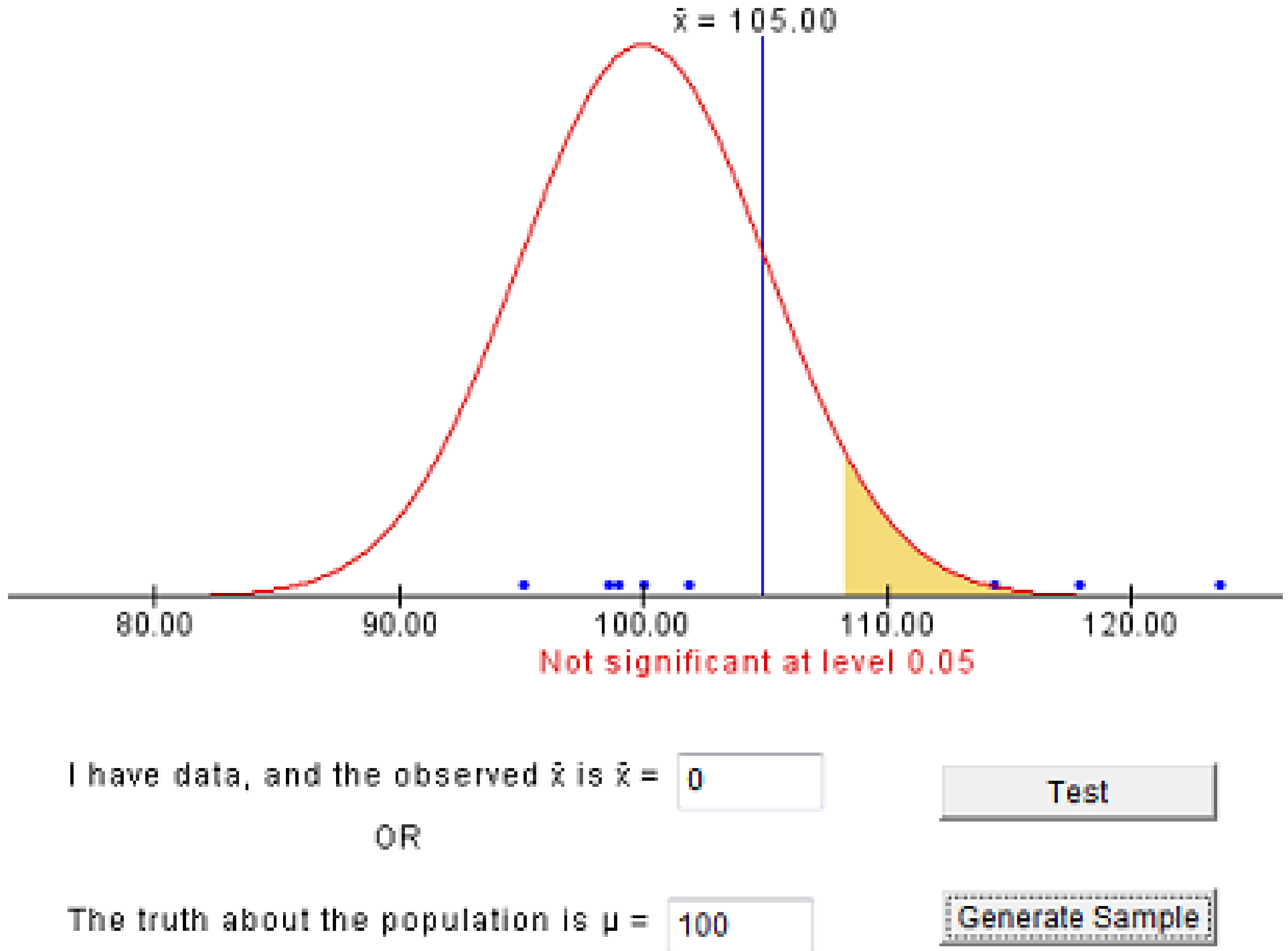
The probability of a type I error is the probability that we reject the null hypothesis given the null hypothesis is true.

- This is easy to calculate as the null distribution will be known.

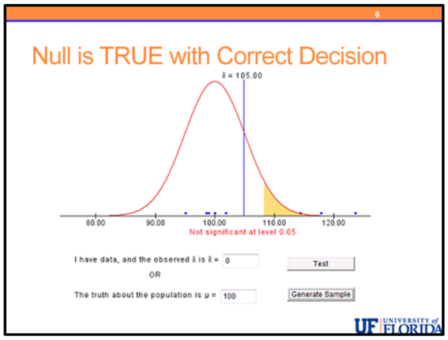
The probability of a type II error is the probability that we fail to reject the null hypothesis given the null hypothesis is false.

- This is impossible to calculate as we do not know the true value if the null hypothesis is false.
- We can only make guesses at the truth and determine this probability in those cases.

Null is TRUE with Correct Decision



UF UNIVERSITY of FLORIDA



Here is a situation where the null hypothesis is that the population mean mu is 100 and the true value is indeed 100. So the null hypothesis is in fact true.

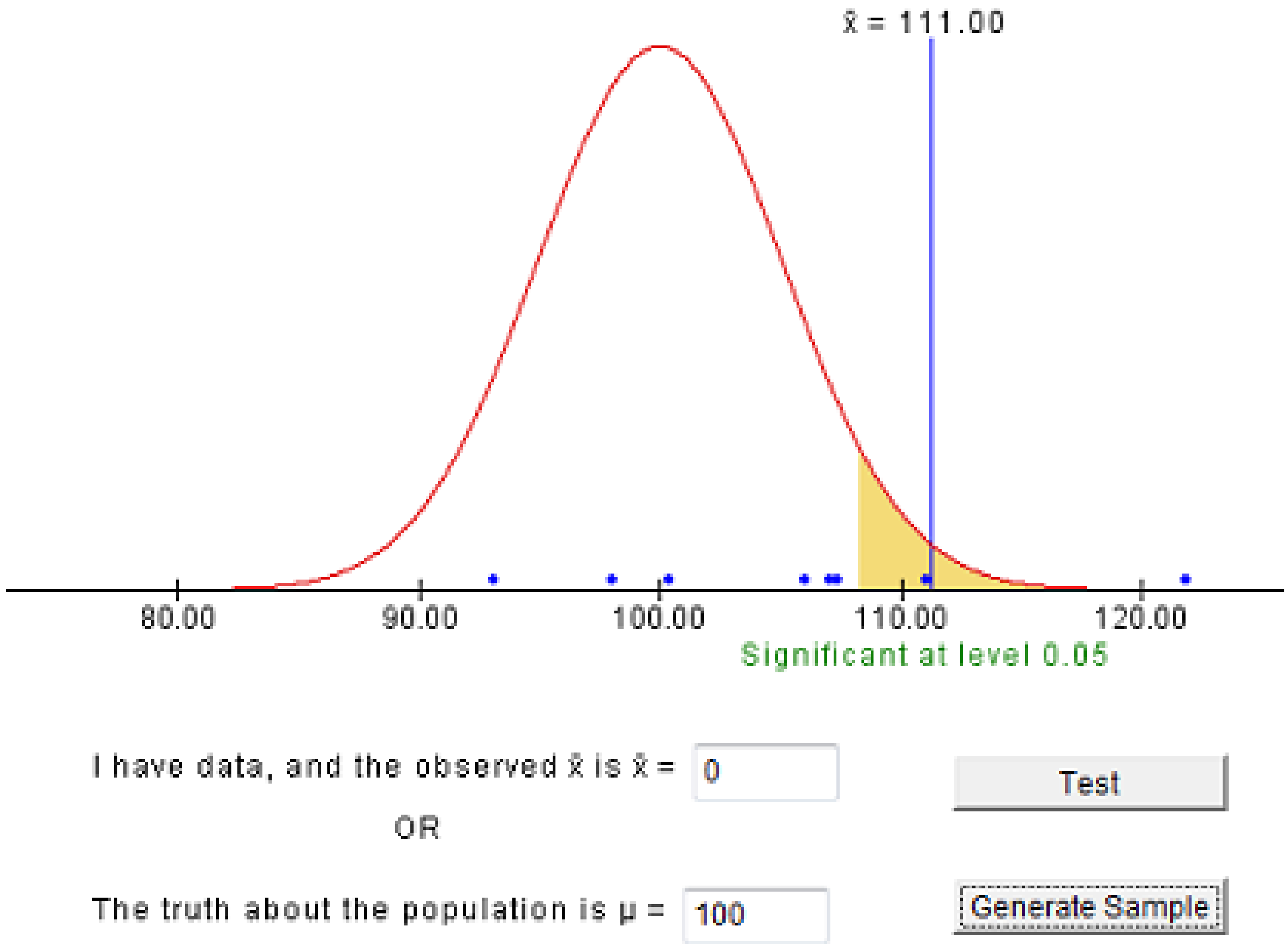
Values in the shaded region would reject the null hypothesis and values not in the shaded region will fail to reject the null.

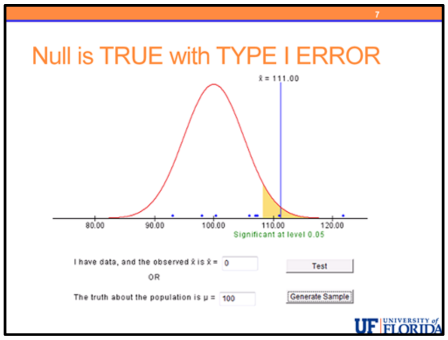
In this case, the data observed did not fall in the shaded region and thus we fail to reject the null.

Since the truth about the population is that the mean is 100, we have made a correct decision.

This should happen about 95% of the time if we use an alpha of 0.05.

Null is TRUE with TYPE I ERROR



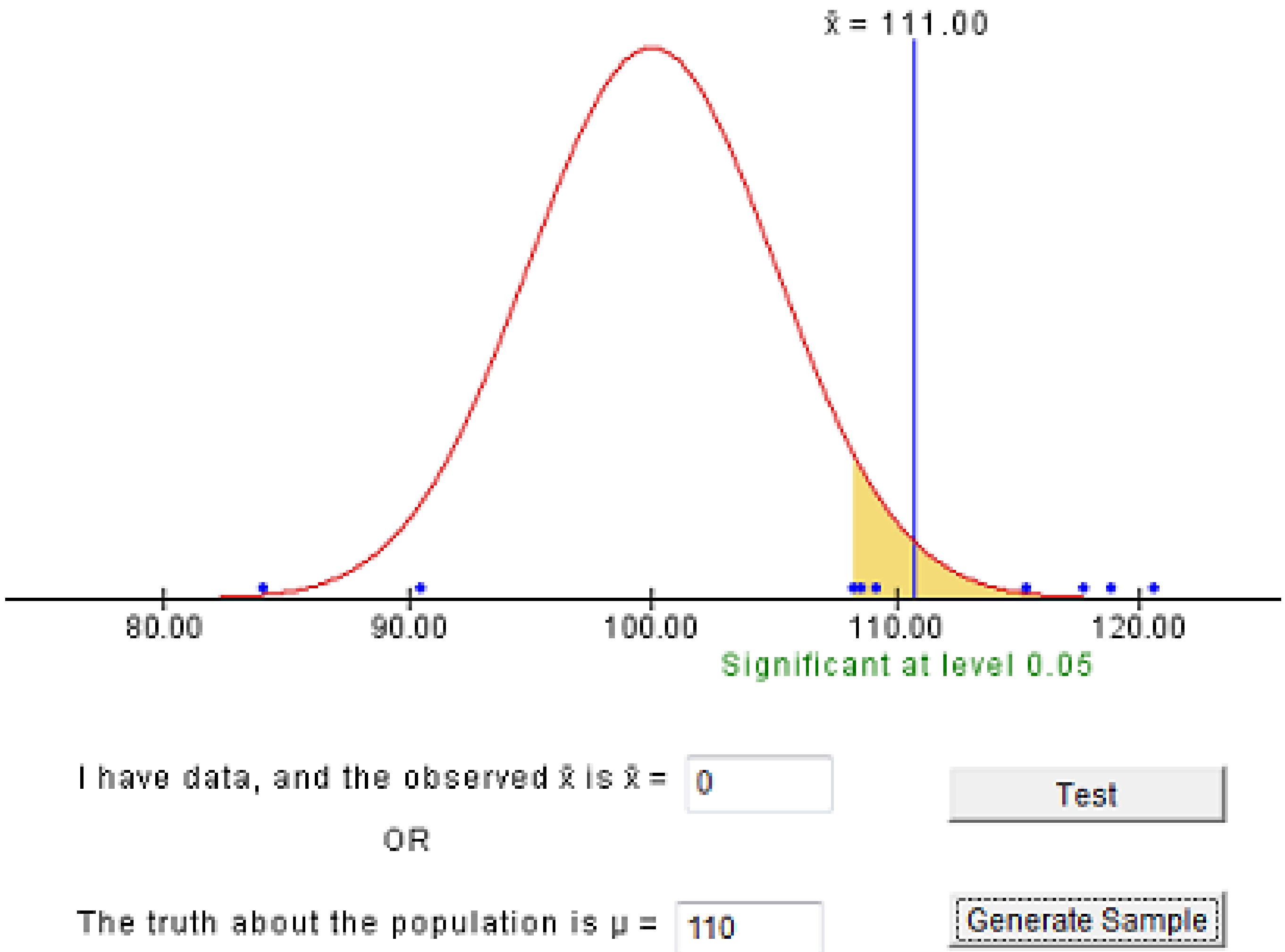


Here again the null hypothesis is that the population mean mu is 100 and the true value is indeed 100. So the null hypothesis is in fact true.

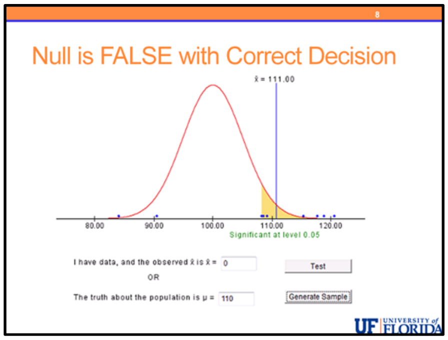
It took a while to find one but here is a sample where the value falls in the shaded region and we reject the null hypothesis when in fact the null hypothesis is true which is a Type I error.

This should happen only about 5% of the time if we use an alpha of 0.05.

Null is FALSE with Correct Decision



UF UNIVERSITY of FLORIDA



Here is a situation where the null hypothesis is that the population mean mu is 100 and the true value is INSTEAD 110. So the null hypothesis is in fact false.

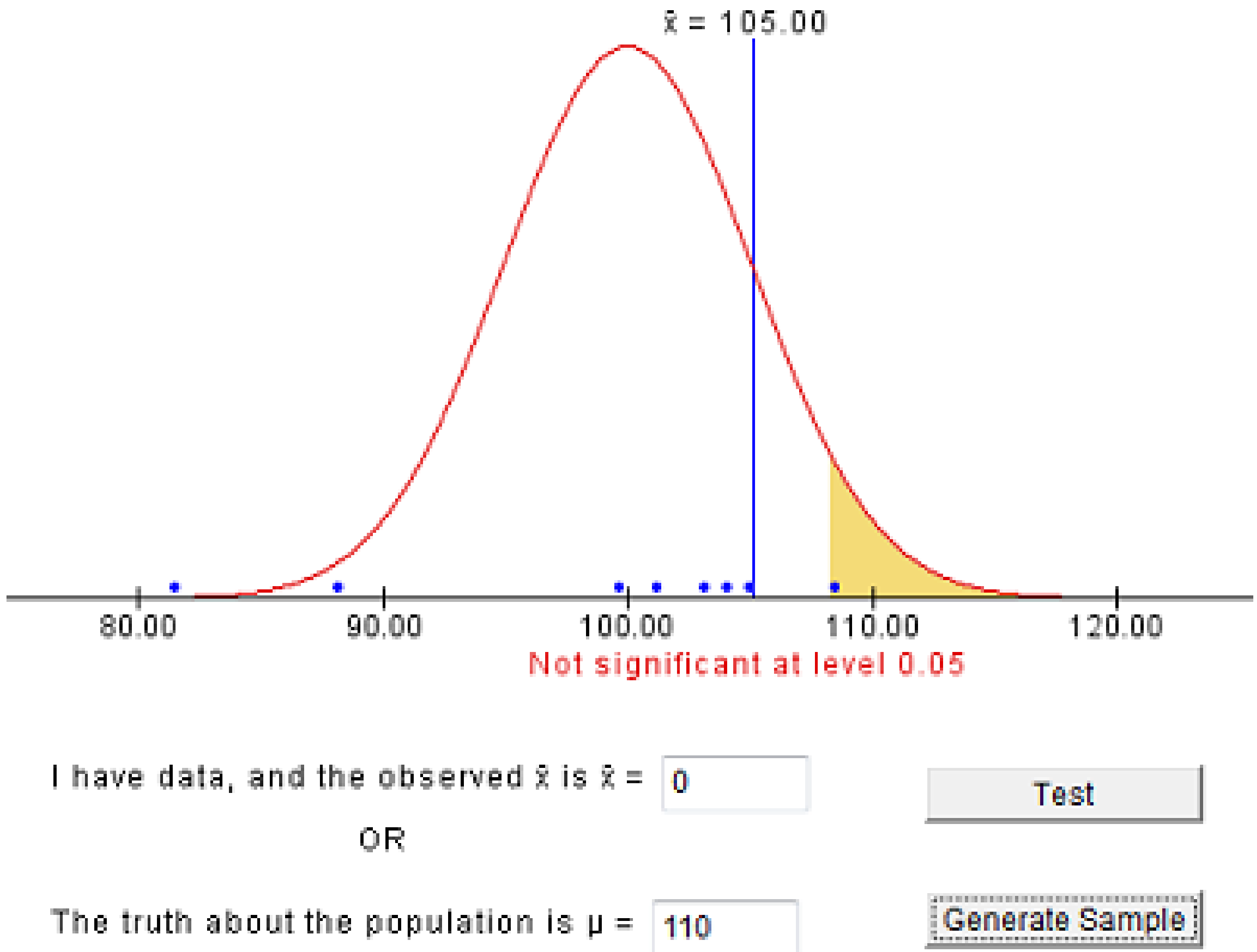
In this case, the data observed did fall in the shaded region and thus we reject the null.

Since the truth about the population is that the mean is 110, we have made a correct decision.

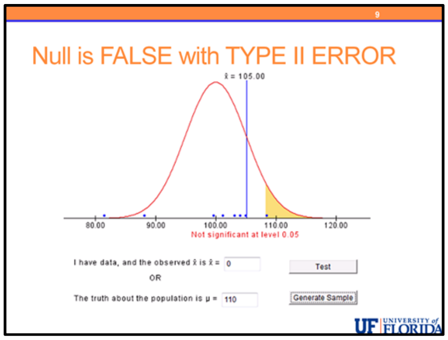
The chance of this happening varies and is based upon the sample size taken and how far the true value is from the null value – which determines how much overlap there is in the underlying sampling distributions. In this case it didn't take long to find a value that rejected the null hypothesis.

We will define this probability as the POWER of the test to detect the specified difference, in this case 10 units.

Null is FALSE with TYPE II ERROR



UF UNIVERSITY of FLORIDA



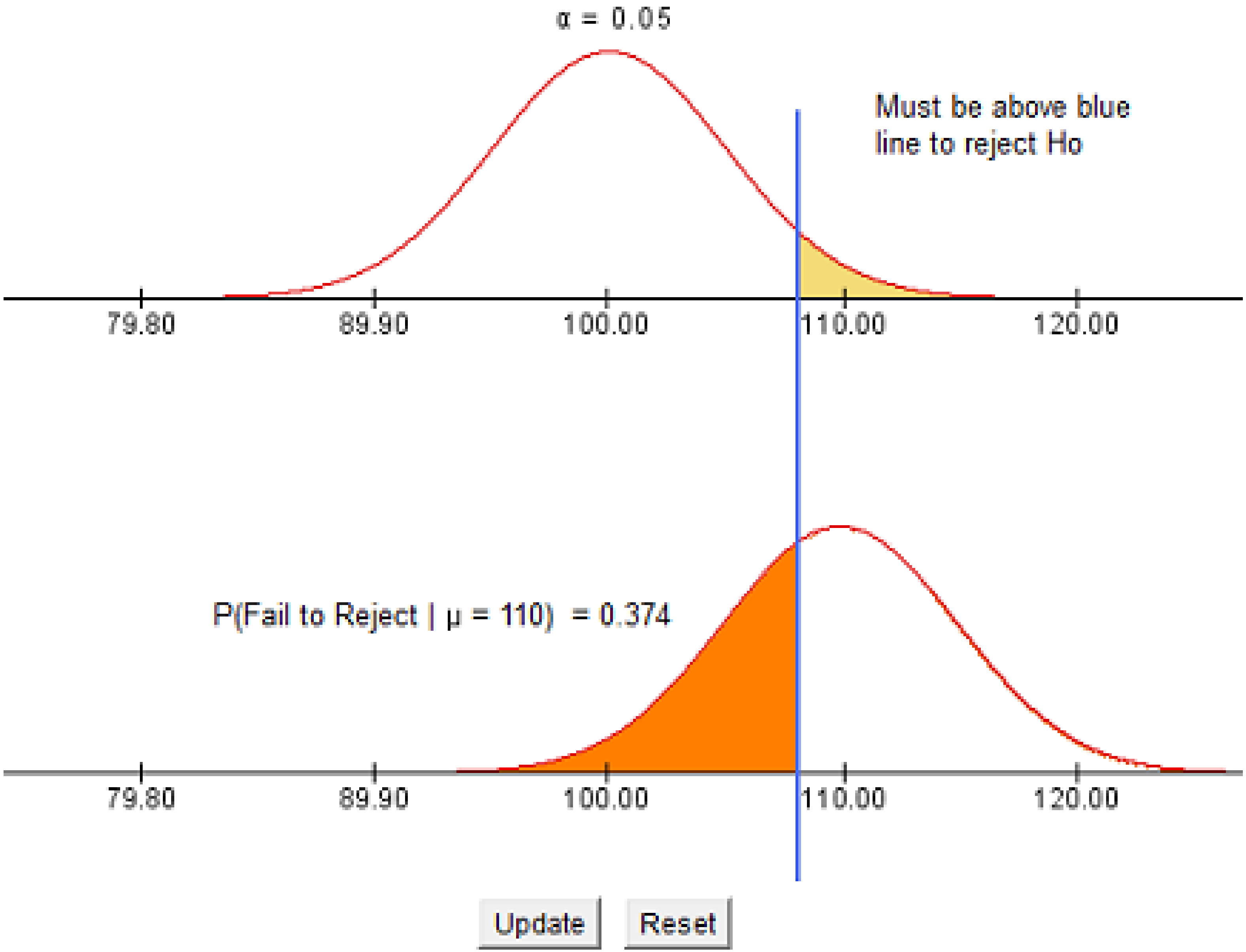
Here again the null hypothesis is that the population mean mu is 100 and the true value is INSTEAD 110. So the null hypothesis is in fact false.

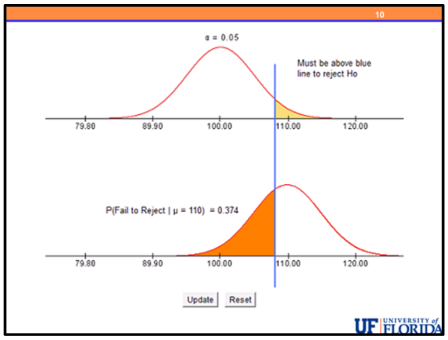
Here is a sample where the value falls outside of the shaded region and we fail to reject the null hypothesis when in fact the null hypothesis is false which is a Type II error.

As in the previous example, the chance of this happening varies and is based upon the sample size taken and how far the true value is from the null value – which determines how much overlap there is in the underlying sampling distributions.

In this case it didn't take long to find a value that failed to reject the null hypothesis.

- This is because the probability of a Type II error in this exact situation is 0.374.
- There is a 37.4% chance of making a type II error if the true mean is 110 and the null hypothesis value is 100 using this sample size and our alpha of 0.05.


α = 0.05
Must be above blue line to reject Ho
79.80
89.90
100.00
110.00
120.00
P(Fail to Reject | μ = 110) = 0.374
79.80
89.90
100.00
110.00
120.00
Update
Reset



Here is an illustration of the Type II error probability.

On the top, we have the distribution assuming the null value of 100 is true.

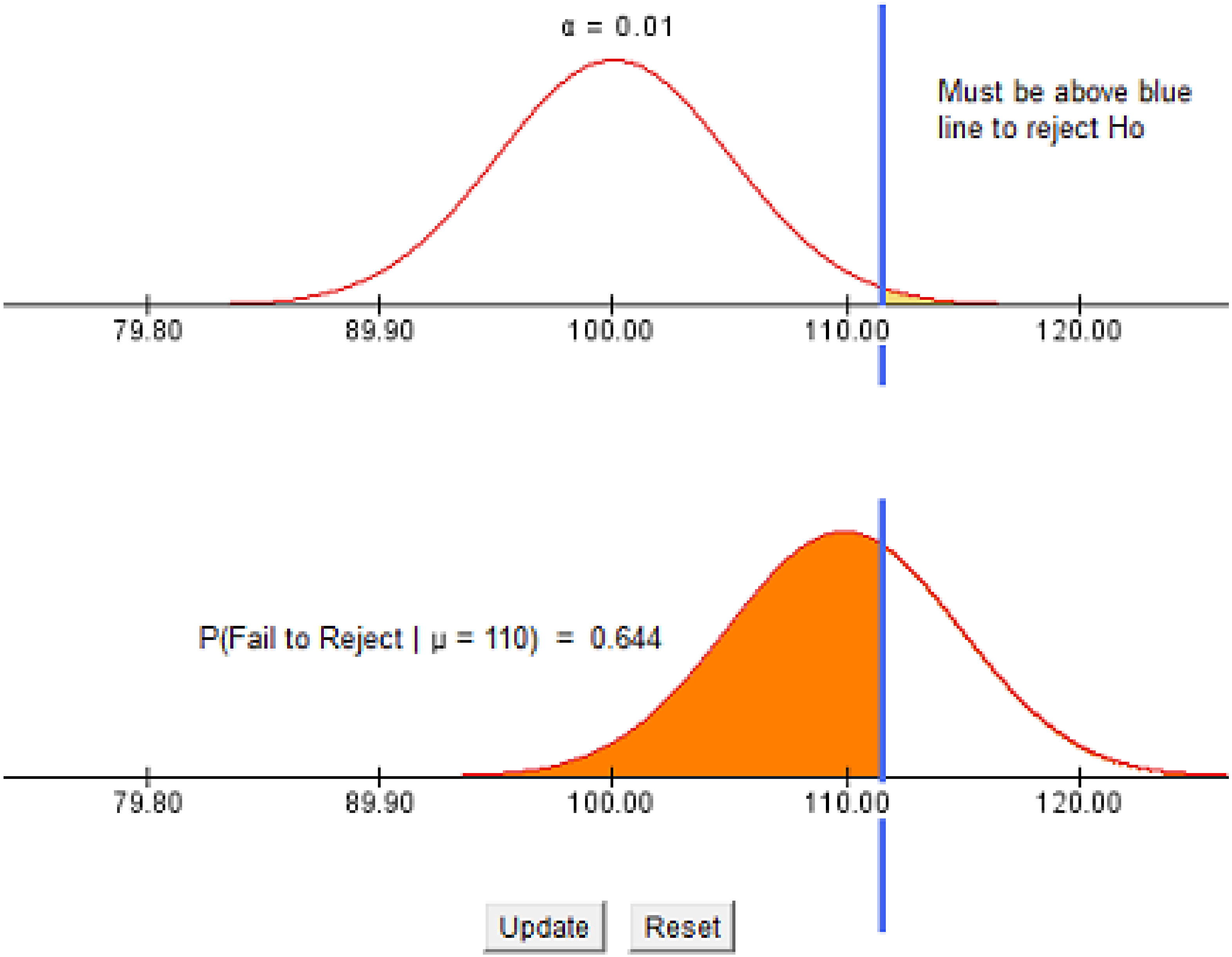
- The blue line represents the cutoff for significance.
- Values above the blue line will result in rejecting the null hypothesis.

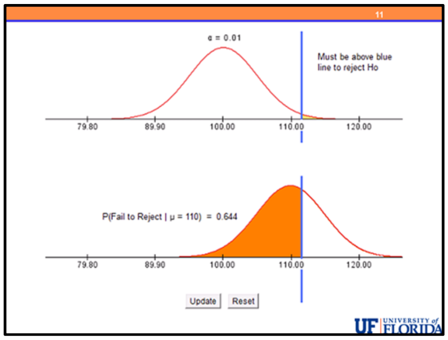
On the bottom, we have the distribution using the true value of 110.

- The blue line continues through.
- If we want to find the probability of failing to reject the null hypothesis – which would result in a Type II error – then we need to find the area to the left of the blue line on this bottom distribution.

You can hopefully believe from the picture that the shaded area in the bottom picture is 37.4%.

You don't need to worry about how to find this value. We simply wanted to help illustrate the idea of the probability of a Type II error.


α = 0.01
Must be above blue line to reject Ho
79.80
89.90
100.00
110.00
120.00
P(Fail to Reject | μ = 110) = 0.644
79.80
89.90
100.00
110.00
120.00
Update
Reset



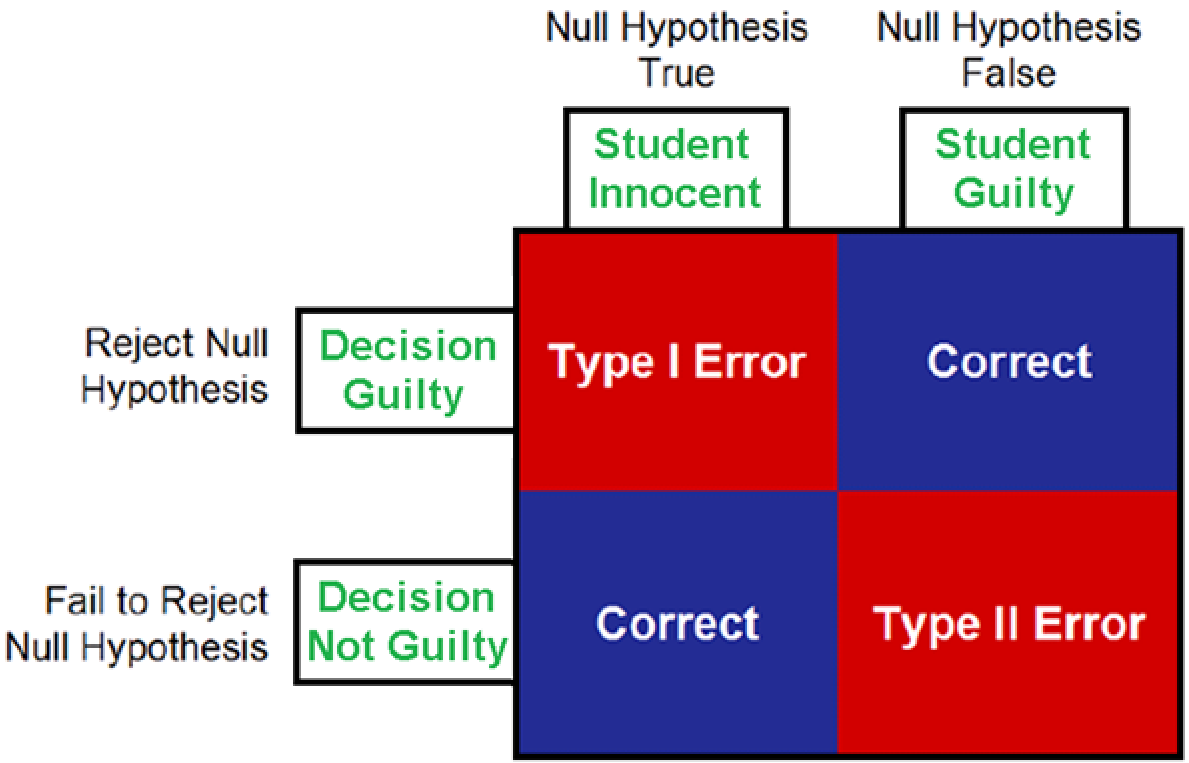
If we change our alpha level, this does have an inverse relationship with the probability of a Type II error.

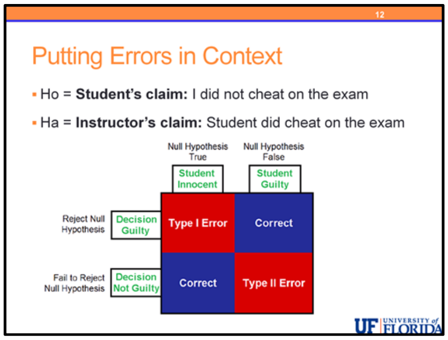
Here we have the same situation but using an alpha level of 0.01. The blue line must move to the right to decrease the chance of a type I error but the result is that there is a much larger chance of making a Type II error. Now it is 64.4%.

What would happen of we reduced the alpha level to 0.10?

Putting Errors in Context

- Ho = **Student's claim:** I did not cheat on the exam
- Ha = **Instructor's claim:** Student did cheat on the exam





Sometimes we are asked to write what a particular type of error or conclusion means in context.

Let's look at our "cheating" example.

Here our null hypothesis is that the student did not cheat on the exam and the alternative hypothesis is that the student did cheat on the exam.

There are four possible outcomes of this process.

There are two possible correct decisions:
- The student did cheat on the exam and the instructor brings enough evidence to reject Ho and conclude the student did cheat on the exam. This is a CORRECT decision!

- The student did not cheat on the exam and the instructor fails to provide enough evidence that the student did cheat on the exam. This is a CORRECT decision!

Both the correct decisions and the possible errors are fairly easy to understand but with the errors, you must be careful to identify and define the two types correctly.

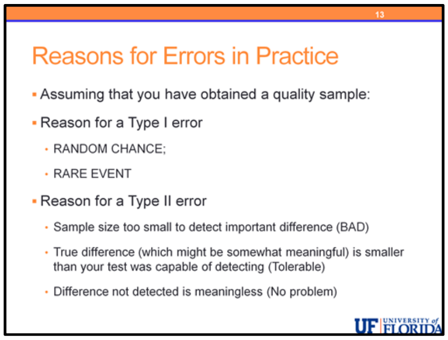
TYPE I Error: Reject Ho when Ho is True
- The student did not cheat on the exam but the instructor brings enough evidence to reject Ho and conclude the student cheated on the exam. This is a Type I Error.

TYPE II Error: Fail to Reject Ho when Ho is False
- The student did cheat on the exam but the instructor fails to provide enough evidence that the student cheated on the exam. This is a Type II Error.

In most situations, including this one, it is more "acceptable" to have a Type II error than a Type I error.
- Although allowing a student who cheats to go unpunished might be considered a very bad problem, punishing a student for something he or she did not do is usually considered to be a more severe error.
- This is one reason we control for our Type I error in the process of hypothesis testing.

Reasons for Errors in Practice

- Assuming that you have obtained a quality sample:
- Reason for a Type I error
  - RANDOM CHANCE;
  - RARE EVENT
- Reason for a Type II error
  - Sample size too small to detect important difference (BAD)
  - True difference (which might be somewhat meaningful) is smaller than your test was capable of detecting (Tolerable)
  - Difference not detected is meaningless (No problem)



Assuming that you have obtained a quality sample:

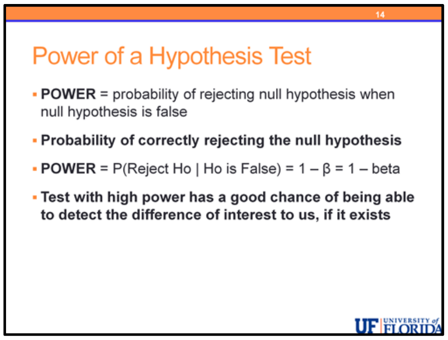
- The reason for a Type I error is random chance.
- When a Type I error occurs, our observed data represented a rare event which indicated evidence in favor of the alternative hypothesis even though the null hypothesis was actually true.

Reason for a Type II error

- The sample size is too small to detect an important difference. This is the worst case, you should have obtained a larger sample.
- The sample size is reasonable for the important difference but the true difference (which might be somewhat meaningful or interesting) is smaller than your test was capable of detecting. This is tolerable as you were not interested in being able to detect this difference when you began your study.
- The sample size is more than adequate, the difference that was not detected is meaningless in practice. This is not a problem at all and is in effect a "correct decision" since the difference you did not detect would have no practical meaning.

Power of a Hypothesis Test

- **POWER** = probability of rejecting null hypothesis when null hypothesis is false
- **Probability of correctly rejecting the null hypothesis**
- **POWER** = P(Reject Ho | Ho is False) = $1 - \beta$ = 1 – beta
- **Test with high power has a good chance of being able to detect the difference of interest to us, if it exists**



The POWER of a hypothesis test is the probability of rejecting the null hypothesis when the null hypothesis is false.

This can also be stated as the probability of correctly rejecting the null hypothesis.

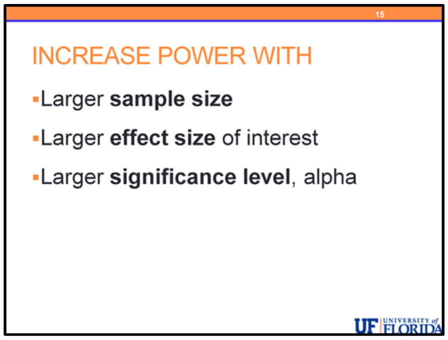
POWER = P(Reject Ho | Ho is False) = $1 - \beta$ = 1 – beta

Power is the test's ability to correctly reject the null hypothesis.

A test with high power has a good chance of being able to detect the difference of interest to us, if it exists.

INCREASE POWER WITH

- Larger **sample size**
- Larger **effect size** of interest
- Larger **significance level**, alpha



Assume that the null hypothesis is false for a given hypothesis test. All else being equal, we have the following:

- Larger samples result in a greater chance to reject the null hypothesis which means an increase in the power of the hypothesis test.

- If the **effect size** is larger, it will become easier for us to detect. This results in a greater chance to reject the null hypothesis which means an increase in the power of the hypothesis test. The effect size varies for each test and is usually closely related to the difference between the hypothesized value and the true value of the parameter under study.

- From the relationship between the probability of a Type I and a Type II error (as α (alpha) decreases, β (beta) increases), we can see that as α (alpha) decreases, Power = $1 - \beta$ = 1 – beta also decreases.

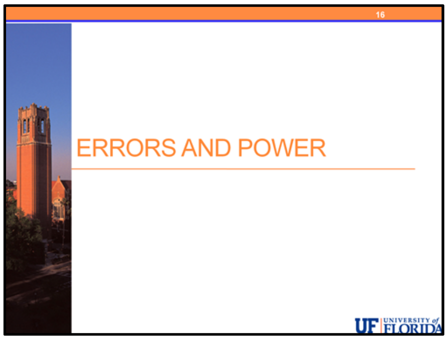
- There are other mathematical ways to change the power of a hypothesis test, such as changing the population standard deviation; however, these are not quantities that we can usually control so we will not discuss them here.

In practice, we specify a significance level and a desired power to detect a difference which will have practical meaning to us and this determines the sample size required for the experiment or study.

For most grants involving statistical analysis, power calculations must be completed to illustrate that the study will have a reasonable chance to detect an important effect. Otherwise, the money spent on the study could be wasted. The goal is usually to have a power close to 80%.

For example, if there is only a 5% chance to detect an important difference between two treatments in a clinical trial, this would result in a waste of time, effort, and money on the study since, when the alternative hypothesis is true, the chance a treatment effect can be found is very small.

ERRORS AND POWER



These concepts are among the most difficult we have discussed this semester.

We will be working with a number of specific hypothesis testing situations for the remainder of the semester and we will be using these tests to help us further explain and grasp the overall process of hypothesis testing as well as the more difficult concepts of errors and power.